Part-Aware and Robust Deep Learning Models for Vision Applications

Jiaxu Miao

Supervisor: Prof. Yi Yang

Faculty of Engineering and Information Technology

University of Technology Sydney

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

June 2021

Declaration

I, Jiaxu Miao declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: 

Jiaxu Miao
June 2021

Acknowledgements

First, I would like to thank my supervisor, Professor Yi Yang. I could not start my doctoral career without his support and help, and it is luckiest for me to pursue a PhD degree under his supervision. I am extremely grateful for his patient guidance, encouragement and selfless support during my doctoral life. He taught me how to start scientific researches and provided helpful advice about my research career. Whenever I met difficulties about academic or personal life, I can get instant help from him.

I would like to thank Dr. Yunchao Wei. He helped me a lot about my research and shared many valuable ideas with me. He also helped me to revise and promote the writings of my papers.

I would like to thank my friend Yu Wu for his help and kindness from the undergraduate period to the doctoral period. I would also like to thank my group members and colleagues, Linchao Zhu, Xin Yu, Pingbo Pan, Ping Liu, Xiaohan Wang, Peike Li, Qianyu Feng, Yutian Lin, Liang Zheng, Xiaojun Chang, Fan Ma, Zhedong Zheng, Hehe Fan, Xuanyi Dong, Yanbin Liu, Hu Zhang, Guang Li, Zhun Zhong, Yawei Luo, Qingji Guan, Guangrui Li, Qi Rao, Ruijie Quan, Tianqi Tang, Yang He, Zongxin Yang, Chen Liang, Xiaolin Zhang, Yuhang Ding, Yunqiu Xu, Youjiang Xu, and many others. Discussions about the research topics with them help me a lot.

I would like to thank my parents, Shiying Miao and Xiufen Wang for their selfless love. They gave me any support when I needed it. They gave me strength and courage when I faced difficulties and challenges in my life.

I would like to thank my beloved wife, Danwen Sun for her sweet love. She is my best friend, my soulmate, and my heart. She always encourages me with her mild sound whenever I feel upset. She taught me to keep an optimistic attitude to life. Thanks for her patience and tremendous help.

Abstract

With the development of deep learning, the deep models based on neural networks play an important role in vision applications. This dissertation focuses on two limitations of previous deep models. First, early approaches for vision tasks usually focus on global representations, while ignoring the discriminative partial features. However, partial representations provide sufficient recognition information for vision tasks and need to be well developed. Second, deep learning models are eager for massive data with labels, which is hard to acquire. The lack of labeled data inherently introduces uncertainty in deep models. Thus, a robust model should not only provide accurate predictions but also estimate uncertainty precisely.

This dissertation presents part-aware and robust deep models for some important vision applications, *i.e.*, the occluded person re-identification (re-id), the interactive video object segmentation (VOS) and the few-shot image classification. Concretely, for the occluded person re-id task, partial features are learned by partitioning the global feature map extracted by neural networks. Pose keypoints are adopted to indicate the visible and occluded parts. The information of occluded parts is depressed. For the interactive VOS, the partial similarity between adjacent frames is important to propagate segmented masks from the previous frame to the current processing frame. Thus, the pixel distances in a local part and the global map are computed for generating masks. For the few-shot image classification, a metric-based Bayesian framework is proposed for generating robust representations and reasoning about uncertainty, including calibration, recognition of out-of-distribution images and robustness against attacks.

In sum, I investigate the significance of the discriminative and robust partial representations and the ability of estimating uncertainty for the deep learning models, and apply them to some common vision applications to illustrate the effectiveness of the deep models.

Table of contents

List of figures